

# INDIGO-DataCloud

Davide Salomoni, INFN-CNAF  
([davide.salomoni@cnaif.infn.it](mailto:davide.salomoni@cnaif.infn.it))

Riunione CCR  
30/3/2015

# INDIGO-DataCloud

- Progetto Horizon 2020 approvato a gennaio 2015 sul bando EINFRA-1.
- **Scopo: sviluppare una piattaforma Cloud open source per dati e calcolo orientata a comunità scientifiche**, installabile su hardware eterogeneo, reperito su infrastrutture Cloud pubbliche o private.
- Il Consorzio di progetto è composto da 26 partner, dislocati in 11 nazioni.
- Forte multi-disciplinarietà: fisica, astrofisica, biologia strutturale, scienze della terra, scienze della vita, bioinformatica, scienze umane, beni culturali



- Partecipazione di 4 grandi industrie
- Finanziamento di 11.1M€ su 30 mesi
- L'ente coordinatore del progetto è l'INFN
- Partecipano le sezioni di Torino, Padova, Bari, Catania e CNAF
- Vedi anche mia comunicazione al WS CCR di dicembre 2014.

# EINFRA-1-2014: Items 4-5 (specifically addressed by INDIGO)

- **Item 4:**

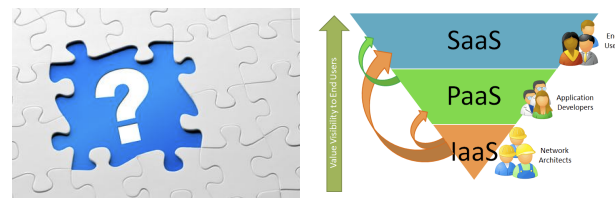
- Large scale virtualization of data/compute centre resources to achieve on-demand compute capacities, improve flexibility for data analysis and avoid unnecessary costly large data transfers.

- **Item 5:**

- Development and adoption of a standards-based computing platform (with open software stack) that can be deployed on different hardware and e-infrastructures (such as clouds providing infrastructure-as-a-service (IaaS), HPC, grid infrastructures...) to abstract application development and execution from available (possibly remote) computing systems. This platform should be capable of federating multiple commercial and/or public cloud resources or services and deliver Platform-as-a-Service (PaaS) adapted to the scientific community with a short learning curve. Adequate coordination and interoperability with existing e-infrastructures (including GÉANT, EGI, PRACE and others) is recommended.

# What is Missing? (1)

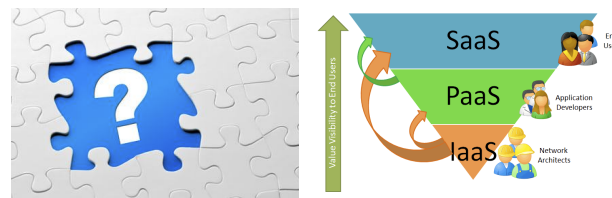
## Or, How to...



- ... **orchestrate and federate** Cloud, Grid and HPC [public or private] resources?
- ... **avoid software and vendor lock-in?**
- ... **overcome performance issues** limiting massive adoption of virtualized Cloud resources in large data centers?
- ... **exploit specialized hardware**, such as GPUs or low-latency interconnections?
- ... **manage dynamic and complex workflows** for scientific data analysis?
- ... **combine data** from multiple sources and stored in multiple locations through incompatible technologies?

# What is Missing? (2)

## Or, How to...



- ... support **federated identities** and provide **privacy** and **distributed authorization** in open Cloud platforms?
- ... provide APIs to exploit the above and **write applications, customizable portals and mobile views?**
- ... **overcome barriers that limit the adoption of PaaS solutions**, such as the use of custom, non-interoperable interfaces, and the limited availability of APIs for technology-independent storage access?
- ... **move beyond static allocation and partitioning** of both storage and computing resources in data centers?
- ... **distribute and deploy applications** in a flexible way?
- ... **exploit** distributed computing and storage resources through **transparent network interconnections**.

# Stato del progetto

- Coordinatore: Davide Salomoni (CNAF)
- Project Legal Signatory: Gaetano Maron (CNAF)
- Project Financial Officer: Luisa Iacono (Padova)
  
- Il Grant Agreement dovrebbe essere nelle fasi finali di approvazione da parte della EC (la firma è attesa entro metà aprile).
- Il Consortium Agreement è in fase avanzata di definizione.
- Inizio formale del progetto: 1/4/2015
- Kick-off Meeting: 22-24/4/2015 (vedi più avanti)

# Riassunto del budget

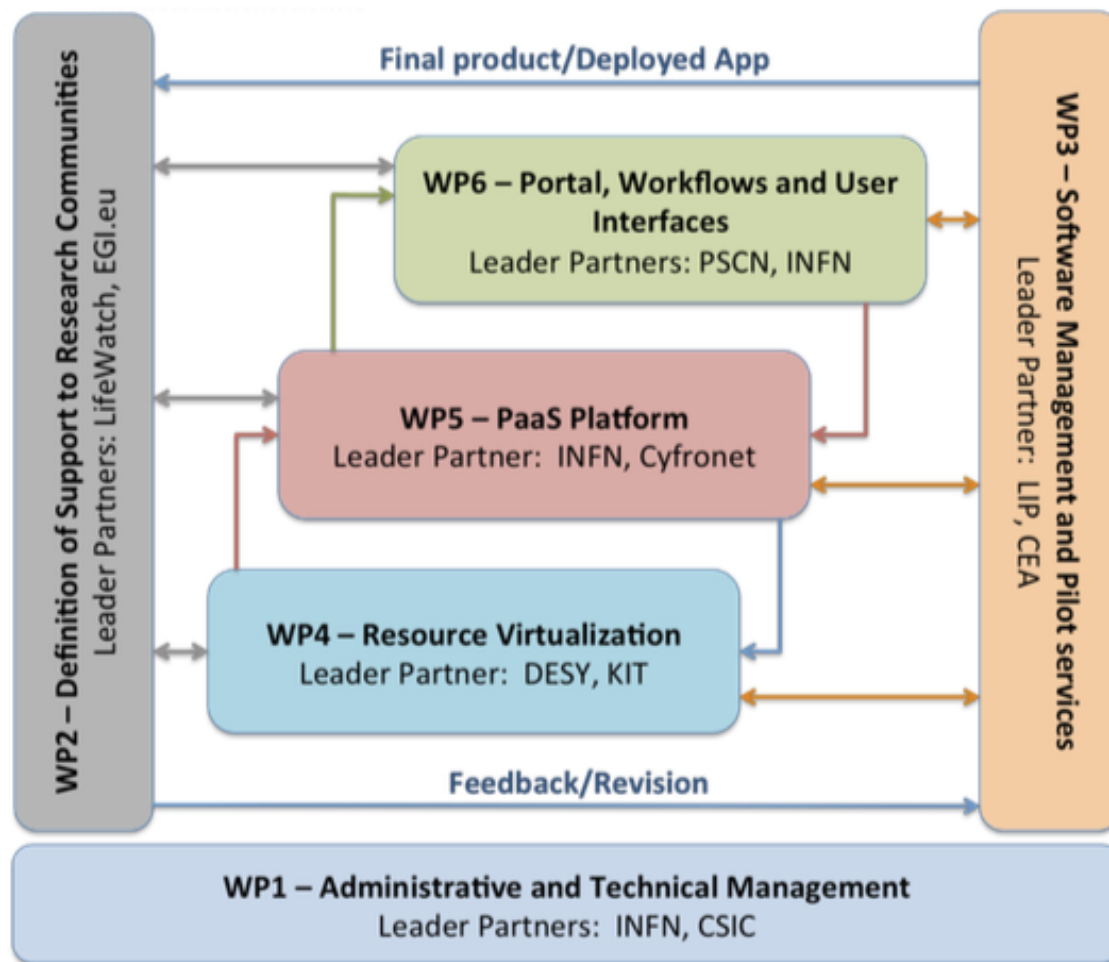
- Budget totale: €11.138.114,00
- Budget INFN: €2.080.614 (circa 18.7%), dei quali:
  - €1.449.000,00 di costi diretti per personale
  - €146.000,00 di altri costi diretti come viaggi, dissemination activities, External Advisory Board
  - €86.864,00 di costi di subcontracting per communication e audit
  - €398.750,00 di costi indiretti (overhead). Di questi costi indiretti l'INFN centrale tratterrà i 2/3, lasciando al progetto circa €132.917,00.
- Totale di 1580 PM (person/month), dei quali 315 all'INFN (circa il 20%)

# Riassunto delle attività

- Le attività sono divise in 6 Work Packages (WP)
  - WP1: Project management, 65PM (4.1%)
  - WP2: Definition of support to Research Communities, 262PM (16.6%)
  - WP3: Software management and pilot services, 214PM (13.5%)
  - WP4: Resource virtualization (compute, storage, network), 342PM (21.6%)
  - WP5: PaaS platform development, 560PM (35.4%)
  - WP6: Science gateways, workflows and toolkits, 137PM (8.7%)



# INDIGO Work Packages



**Figure 4: Diagram showing the interrelation among the Work Packages**

# L'INFN in INDIGO

- Coordinamento globale: D.Salomoni (PI), L.Iacono (FO)
- Technical Director: G.Donvito
- Coordinamento del WP5: G.Donvito
- Co-coordinamento del WP6: R.Barbera
- Partecipazione ai singoli WP:
  - 42 PM in WP3 (19.6%, 11 partners) – Software management and pilot services
  - 69 PM in WP4 (20.2%, 12 partners) – Resource virtualization
  - 139 PM in WP5 (24.8%, 15 partners) – PaaS platform development
  - 45 PM in WP6 (32.8%, 5 partners) – Science gateways, workflows and toolkits

# Attività nelle sezioni finanziate

- I contributi delle sezioni INFN finanziate dal progetto sono così suddivisi:
  - Torino: WP1, WP3.3, WP4.1, WP4.3, WP5.1, WP5.3
  - Padova: WP1, WP3.3, WP4.1, WP5.3
  - Bari: WP3.3, WP4.2, WP5.1, WP5.3, WP5.4
  - Catania: WP3.3, WP5.2, WP5.4, WP6.1, WP6.2
  - CNAF: WP1, WP3.2, WP3.3, WP4.1, WP4.2, WP4.3, WP5.1, WP5.2, WP5.4, WP6.2

# Contratti a tempo determinato

- Sono inizialmente previste le seguenti assunzioni, da realizzarsi integralmente o parzialmente su fondi di progetto e di durata tra i 24 e i 30 mesi:
  - 11 FTE con contratto art.36, così suddivisi:
    - 1 posizione a Torino
    - 2 posizioni a Padova
    - 2 posizioni a Bari
    - 2 posizioni a Catania
    - 4 posizioni a CNAF
  - 1 FTE con contratto art.15 per attività di gestione amministrativa a Padova

# Partecipazione di personale strutturato

- Ogni sezione partecipante al progetto ha ***preliminariamente*** dichiarato quale personale strutturato si prevede che partecipi, su quali attività e per quanto tempo (dettaglio disponibile su richiesta). In sintesi:
  - Torino: 15PM
  - Padova: 12PM + amministrazione
  - Bari: 17PM
  - Catania: 9PM (+ 4PM unfunded)
  - CNAF: 54PM (+ altri PM da personale su fondi interni)Per un totale *preliminare* di circa 100-140PM.

# Impatto sugli esperimenti INFN

- Per garantire che gli esperimenti di interesse INFN siano adeguatamente rappresentati in INDIGO, vorremmo operare perché:
  - **A livello di progetto**, si inserisca un rappresentante di esperimenti INFN nel Technical Board.
    - Dalla bozza attuale di CA: “The TB may also include representatives for other infrastructures or projects relevant for INDIGO, upon proposal by the PMB.” – ad esempio, come rappresentante di esperimenti LHC
  - **A livello interno INFN**, vi siano operativamente uno o più rappresentanti di esperimenti INFN (**che abbiano tempo, competenza e volontà concreta di partecipare**) nelle attività di integrazione e test (3.3), sviluppo compute, storage & network virtualization (4.1, 4.2, 4.3) e architettura PaaS (5.1).
    - Abbiamo preliminarmente contattato per questo compito Tommaso Boccali, che dovrebbe operare come punto di contatto tra attività di esperimento legate al calcolo INFN ed INDIGO.
    - **Ogni feedback relativo al programma tecnico di INDIGO è sempre benvenuto!**

# Relazione con CCR

- INDIGO è un progetto che nasce e cresce su esperienze direttamente connesse ad attività discusse e gestite in CCR. Tra queste, in particolare, il gruppo Cloud INFN, ma anche progetti legati ad esempio ad autenticazione, security o gestione della rete.
- Pensiamo che sia auspicabile e necessario continuare questa attività di collaborazione tecnica diretta con CCR, attraverso ad esempio:
  - Comunicazioni periodiche sullo stato delle attività
  - Ove possibile, partecipazione alle attività anche da parte di sezioni non direttamente presenti in INDIGO (con possibilità ovviamente di contribuire anche a paper, conferenze, etc.)

# Richieste di budget a CCR

- Il progetto prevede alcuni fondi di missione. Tuttavia, vi sono missioni (e in generale spese) che, pur importanti per il progetto, non sono direttamente rendicontabili su di esso; per esempio, trasferte presso la EC, partecipazioni ad eventi come EGI technical/user forums, SuperComputing, GDB.
- Per questo motivo abbiamo richiesto a CCR una integrazione di fondi di missione per il 2015 (vedi referaggio domani).



# Kick-off Meeting

- Il Kick-off Meeting di INDIGO si terrà **dal 22 al 24 aprile 2015** presso l'area di ricerca del CNR a Bologna.
- Una agenda preliminare è disponibile su <https://agenda.infn.it/conferenceDisplay.py?ovw=True&confId=9460>
- La partecipazione è gratuita (verranno usati i fondi di progetto) ma la registrazione è obbligatoria.
  - La partecipazione di personale INFN non direttamente connesso al progetto per portare contributi tecnici (è un project kick-off meeting, non un workshop generale) ad es. legati ad attività CCR è possibile ed anzi auspicabile. Chi fosse interessato a partecipare può contattare Claudio, Dario o me.

# GRAZIE!

[Davide.Salomoni@cnaif.infn.it](mailto:Davide.Salomoni@cnaif.infn.it)